



2570  
2500

## RAW SEQUENCE LISTING ERROR REPORT

The Biotechnology Systems Branch of the Scientific and Technical Information Center (STIC) detected errors when processing the following computer readable form:

Application Serial Number: 10/091,244  
Source: CIPR  
Date Processed by STIC 8/15/02

THE ATTACHED PRINTOUT EXPLAINS DETECTED ERRORS.

PLEASE FORWARD THIS INFORMATION TO THE APPLICANT BY EITHER:

- 1) INCLUDING A COPY OF THIS PRINTOUT IN YOUR NEXT COMMUNICATION TO THE APPLICANT, WITH A NOTICE TO COMPLY or,
- 2) TELEPHONING APPLICANT AND FAXING A COPY OF THIS PRINTOUT, WITH A NOTICE TO COMPLY

FOR CRF SUBMISSION QUESTIONS, PLEASE CONTACT MARK SPENCER, 703-308-4212.

FOR SEQUENCE RULES INTERPRETATION, PLEASE CONTACT ROBERT WAX, 703-308-4216.

PATENTIN 2.1 e-mail help: [patin21help@uspto.gov](mailto:patin21help@uspto.gov) or phone 703-306-4119 (R. Wax)

PATENTIN 3.0 e-mail help: [patin3help@uspto.gov](mailto:patin3help@uspto.gov) or phone 703-306-4119 (R. Wax)

TO REDUCE ERRORED SEQUENCE LISTINGS, PLEASE USE THE **CHECKER VERSION 3.1 PROGRAM**, ACCESSIBLE THROUGH THE U.S. PATENT AND TRADEMARK OFFICE WEBSITE. SEE BELOW FOR ADDRESS:

<http://www.uspto.gov/web/offices/pac/checker>

Applicants submitting genetic sequence information electronically on diskette or CD-Rom should be aware that there is a possibility that the disk/CD-Rom may have been affected by treatment given to all incoming mail.

Please consider using alternate methods of submission for the disk/CD-Rom or replacement disk/CD-Rom.

Any reply including a sequence listing in electronic form should NOT be sent to the 20231 zip code address for the United States Patent and Trademark Office, and instead should be sent via the following to the indicated addresses:

1. EFS-Bio (<<http://www.uspto.gov/ebc/efs/downloads/documents.htm>> , EFS Submission User Manual - ePAVE)
2. U.S. Postal Service: U.S. Patent and Trademark Office, Box Sequence, P.O. Box 2327, Arlington, VA 22202
3. Hand Carry directly to:  
U.S. Patent and Trademark Office, Technology Center 1600, Reception Area, 7<sup>th</sup> Floor, Examiner Name, Sequence Information, Crystal Mall One, 1911 South Clark Street, Arlington, VA 22202  
Or  
U.S. Patent and Trademark Office, Box Sequence, Customer Window, Lobby, Room 1B03, Crystal Plaza Two, 2011 South Clark Place, Arlington, VA 22202
4. Federal Express, United Parcel Service, or other delivery service to: U.S. Patent and Trademark Office, Box Sequence, Room 1B03-Mailroom, Crystal Plaza Two, 2011 South Clark Place, Arlington, VA 22202

# Raw Sequence Listing Error Summary

## ERROR DETECTED

## SUGGESTED CORRECTION

SERIAL NUMBER: 16/0011/244

ATTN: NEW RULES CASES: PLEASE DISREGARD ENGLISH "ALPHA" HEADERS, WHICH WERE INSERTED BY PTO SOFTWARE

- 1      Wrapped Nucleics  
    Wrapped Aminos      The number/text at the end of each line "wrapped" down to the next line. This may occur if your file was retrieved in a word processor **after** creating it. Please adjust your right margin to .3; this will prevent "wrapping."
- 2      Invalid Line Length      The rules require that a line **not exceed** 72 characters in length. This includes white spaces.
- 3      Misaligned Amino  
    Numbering      The numbering under each 5<sup>th</sup> amino acid is misaligned. Do **not** use tab codes between numbers; use **space characters**, instead.
- 4      Non-ASCII      The submitted file was **not** saved in ASCII(DOS) text, as **required** by the Sequence Rules. Please **ensure your subsequent submission is saved in ASCII text**.
- 5      Variable Length      Sequence(s)      contain n's or Xaa's representing more than one residue. **Per Sequence Rules, each n or Xaa can only represent a single residue.** Please present the **maximum** number of each residue having variable length and indicate in the <220>-<223> section that some may be missing.
- 6      PatentIn 2.0  
    "bug"      A "bug" in PatentIn version 2.0 has caused the <220>-<223> section to be missing from amino acid sequences(s)     . Normally, PatentIn would automatically generate this section from the previously coded nucleic acid sequence. Please manually copy the relevant <220>-<223> section to the subsequent amino acid sequence. **This applies to the mandatory <220>-<223> sections for Artificial or Unknown sequences.**
- 7      Skipped Sequences  
    (OLD RULES)      Sequence(s)      missing. If intentional, please insert the following lines for **each** skipped sequence:  
    (2) INFORMATION FOR SEQ ID NO:X: (insert SEQ ID NO where "X" is shown)  
    (i) SEQUENCE CHARACTERISTICS: (Do not insert any subheadings under this heading)  
    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:X: (insert SEQ ID NO where "X" is shown)  
    This sequence is intentionally skipped  
  
    Please also adjust the "(ii) NUMBER OF SEQUENCES:" response to **include** the skipped sequences.
- 8      Skipped Sequences  
    (NEW RULES)      Sequence(s)      missing. If intentional, please insert the following lines for **each** skipped sequence.  
    <210> sequence id number  
    <400> sequence id number  
    000
- 9      Use of n's or Xaa's  
    (NEW RULES)      Use of n's and/or Xaa's have been detected in the Sequence Listing.  
    Per 1.823 of Sequence Rules, use of <220>-<223> is **MANDATORY** if n's or Xaa's are present.  
    In <220> to <223> section, please explain location of **n** or **Xaa**, and which residue **n** or **Xaa** represents.
- 10      Invalid <213>  
    Response      Per 1.823 of Sequence Rules, the only **valid** <213> responses are: Unknown, Artificial Sequence, or scientific name (Genus/species). <220>-<223> section is **required** when <213> response is Unknown or is Artificial Sequence
- 11      Use of <220>      Sequence(s)      missing the <220> "Feature" and associated numeric identifiers and responses.  
    Use of <220> to <223> is **MANDATORY** if <213> "Organism" response is "Artificial Sequence" or "Unknown." Please explain source of genetic material in <220> to <223> section.  
    (See "Federal Register," 06/01/1998, Vol. 63, No. 104, pp. 29631-32) (Sec. 1.823 of Sequence Rules)
- 12      PatentIn 2.0  
    "bug"      Please do not use "Copy to Disk" function of PatentIn version 2.0. This causes a corrupted file, resulting in missing mandatory numeric identifiers and responses (as indicated on raw sequence listing). Instead, please use "File Manager" or any other manual means to copy file to floppy disk.
- 13      Misuse of n      n can only be used to represent a single nucleotide in a nucleic acid sequence. N is not used to represent any value not specifically a nucleotide.



OIPE

## RAW SEQUENCE LISTING

DATE: 08/13/2002

PATENT APPLICATION: US/10/091,244

TIME: 13:46:44

Input Set : N:\jumbos\10091244DC.txt

Output Set: N:\CRF4\08132002\J091244.raw

```

4 <110> APPLICANT: Gokhale, Rajesh
5      Tsuji, Stuart
6      Khosla, Chaitan
7 <120> TITLE OF INVENTION: METHODS TO MEDIATE POLYKETIDE SYNTHASE
8      MODULE EFFECTIVENESS
11 <130> FILE REFERENCE: 300622004620
13 <140> CURRENT APPLICATION NUMBER: US 10/091,244
14 <141> CURRENT FILING DATE: 2002-03-04
16 <150> PRIOR APPLICATION NUMBER: 09/500,747
17 <151> PRIOR FILING DATE: 2000-02-09
19 <150> PRIOR APPLICATION NUMBER: 60/119,363
20 <151> PRIOR FILING DATE: 1999-02-09
22 <150> PRIOR APPLICATION NUMBER: 60/272,985
23 <151> PRIOR FILING DATE: 2001-03-02
25 <150> PRIOR APPLICATION NUMBER: 60/272,987
26 <151> PRIOR FILING DATE: 2001-03-02
28 <160> NUMBER OF SEQ ID NOS: 24
30 <170> SOFTWARE: FastSEQ for Windows Version 4.0
32 <210> SEQ ID NO: 1
33 <211> LENGTH: 15
34 <212> TYPE: DNA
35 <213> ORGANISM: Artificial Sequence
37 <220> FEATURE:
38 <223> OTHER INFORMATION: Nhe site upstream of the KS at position 7570
40 <400> SEQUENCE: 1
41 gctagcgagc cgatc 15
43 <210> SEQ ID NO: 2
44 <211> LENGTH: 15
45 <212> TYPE: DNA
46 <213> ORGANISM: Artificial Sequence
47 <220> FEATURE:
48 <223> OTHER INFORMATION: Nhe site upstream of the KS at position 28710
51 <400> SEQUENCE: 2
52 gctagcgagc cgatc 15
54 <210> SEQ ID NO: 3
55 <211> LENGTH: 18
56 <212> TYPE: PRT
57 <213> ORGANISM: Artificial Sequence
58 <220> FEATURE:
59 <223> OTHER INFORMATION: Intra-polypeptide linker
61 <400> SEQUENCE: 3
63 Gly Gly Ala Thr Gly Ala Glu Gln Ala Ala Pro Ala Thr Thr Ala Pro
64 1 5 10 15

```

## RAW SEQUENCE LISTING

DATE 08/13/2002

PATENT APPLICATION: US/10/091,244

TIME 13:46:44

Input Set : N:\jumbos\10091244DC.txt

Output Set: N:\CRF4\08132002\J091244.raw

```

65 Val Asp
69 <210> SEQ ID NO: 4
70 <211> LENGTH: 18
71 <212> TYPE: PRT
72 <213> ORGANISM: Artificial Sequence
74 <220> FEATURE:
75 <223> OTHER INFORMATION: Intra-polypeptide linker
77 <400> SEQUENCE: 4
78 Val Gly Asp Ala Asp Gln Ala Ala Val Arg Val Val Gly Ala Ala Asp
79 1 5 10 15
81 Glu Ser
84 <210> SEQ ID NO: 5
85 <211> LENGTH: 21
86 <212> TYPE: PRT
87 <213> ORGANISM: Artificial Sequence
89 <220> FEATURE:
90 <223> OTHER INFORMATION: Intra-polypeptide linker
92 <400> SEQUENCE: 5
93 Val Gly Ala Ala Glu Ala Glu Gln Ala Pro Ala Leu Val Arg Glu Val
94 1 5 10 15
95 Pro Lys Asp Ala Asp
96 20
99 <210> SEQ ID NO: 6
100 <211> LENGTH: 17
101 <212> TYPE: PRT
102 <213> ORGANISM: Artificial Sequence
104 <220> FEATURE:
105 <223> OTHER INFORMATION: Intra-polypeptide linker
107 <400> SEQUENCE: 6
108 Phe Gly Ser Ala Ala Asn Arg Pro Ala Glu Ile Gly Thr Ala Ala Ala
109 1 5 10 15
110 Glu
114 <210> SEQ ID NO: 7
115 <211> LENGTH: 17
116 <212> TYPE: PRT
117 <213> ORGANISM: Artificial Sequence
119 <220> FEATURE:
120 <223> OTHER INFORMATION: Intra-polypeptide linker
122 <400> SEQUENCE: 7
123 Leu Gly Glu Arg Pro Ala Ala Pro Ala Pro Val Thr Arg Asp Val Ser
124 1 5 10 15
125 Asp
129 <210> SEQ ID NO: 8
130 <211> LENGTH: 19
131 <212> TYPE: PRT
132 <213> ORGANISM: Artificial Sequence
134 <220> FEATURE:
135 <223> OTHER INFORMATION: Intra-polypeptide linker
137 <400> SEQUENCE: 8

```

## RAW SEQUENCE LISTING

DATE: 08/13/2002

PATENT APPLICATION: US/10/091,244

TIME: 13:46:44

Input Set : N:\jumbos\10091244DC.txt

Output Set: N:\CRF4\08132002\J091244.raw

```

138 Gly Glu Thr Val Ala Gly Ala Pro Ala Thr Pro Val Thr Thr Val Ala
139 1 5 10 15
140 Asp Ala Gly
141 <210> SEQ ID NO: 9
142 <211> LENGTH: 21
143 <212> TYPE: PRT
144 <213> ORGANISM: Artificial Sequence
145 <220> FEATURE:
146 <223> OTHER INFORMATION: Intra-polypeptide linker
147 <400> SEQUENCE: 9
148 Glu Leu Phe Thr Gly Glu Asn Pro Ala Pro Val Arg Gly Pro Val Ser
149 1 5 10 15
150 Ala Val Gly Gln Asp
151 20
152 <210> SEQ ID NO: 10
153 <211> LENGTH: 21
154 <212> TYPE: PRT
155 <213> ORGANISM: Artificial Sequence
156 <220> FEATURE:
157 <223> OTHER INFORMATION: Intra-polypeptide linker
158 <400> SEQUENCE: 10
159 Glu Leu Phe Thr Gly Glu Asn Pro Ala Pro Val Arg Gly Pro Val Ser
160 1 5 10 15
161 Val Val Gly Gln Asp
162 20
163 <210> SEQ ID NO: 11
164 <211> LENGTH: 20
165 <212> TYPE: PRT
166 <213> ORGANISM: Artificial Sequence
167 <220> FEATURE:
168 <223> OTHER INFORMATION: Intra-polypeptide linker
169 <400> SEQUENCE: 11
170 Glu Leu Phe Thr Gly Glu Asn Pro Ala Pro Val Arg Gly Pro Val Ser
171 1 5 10 15
172 Ala Gly Gln Asp
173 20
174 <210> SEQ ID NO: 12
175 <211> LENGTH: 30
176 <212> TYPE: PRT
177 <213> ORGANISM: Artificial Sequence
178 <220> FEATURE:
179 <223> OTHER INFORMATION: N-Terminal Inter-polypeptide linker
180 <400> SEQUENCE: 12
181 Val Thr Asp Ser Glu Lys Val Ala Glu Tyr Leu Arg Arg Ala Thr Leu
182 1 5 10 15
183 Asp Leu Arg Ala Ala Arg Gln Arg Ile Arg Glu Leu Glu Ser
184 20 25 30
185 <210> SEQ ID NO: 13
186 <211> LENGTH: 36

```

## RAW SEQUENCE LISTING

DATE 08/13/2002

PATENT APPLICATION: US/10/091,244

TIME 13:46:44

Input Set : N:\jumbos\10091244DC.txt

Output Set: N:\CRF4\08132002\J091244.raw

```

206 <212> TYPE: PRT
207 <213> ORGANISM: Artificial Sequence
208 <220> FEATURE:
209 <223> OTHER INFORMATION: N-Terminal Inter-polypeptide linker
210 <400> SEQUENCE: 13
211 Met Ser Gly Asp Asn Gly Met Thr Glu Glu Lys Leu Arg Arg Tyr Leu
212 1 5 10 15
213 Lys Arg Thr Val Thr Glu Leu Asp Ser Val Thr Ala Arg Leu Arg Glu
214 20 25 30
215 Val Glu His Arg Ala Gly
216 35
221 <210> SEQ ID NO: 14
222 <211> LENGTH: 34
223 <212> TYPE: PRT
224 <213> ORGANISM: Artificial Sequence
225 <220> FEATURE:
226 <223> OTHER INFORMATION: N-Terminal Inter-polypeptide linker
227 <400> SEQUENCE: 14
228 Met Ser Ala Pro Asn Glu Gln Ile Val Asp Ala Leu Arg Ala Ser Leu
229 1 5 10 15
230 Lys Glu Asn Val Arg Leu Gln Gln Glu Asn Ser Ala Leu Ala Ala
231 20 25 30
232 Ala Ala
233 <210> SEQ ID NO: 15
234 <211> LENGTH: 34
235 <212> TYPE: PRT
236 <213> ORGANISM: Artificial Sequence
237 <220> FEATURE:
238 <223> OTHER INFORMATION: N-Terminal Inter-polypeptide linker
239 <400> SEQUENCE: 15
240 Val Ser Ala Ser Tyr Glu Lys Val Val Glu Ala Leu Arg Lys Ser Leu
241 1 5 10 15
242 Glu Glu Val Gly Thr Leu Lys Lys Arg Asn Arg Gln Leu Ala Asp Ala
243 20 25 30
244 Ala Gly
245 <210> SEQ ID NO: 16
246 <211> LENGTH: 33
247 <212> TYPE: PRT
248 <213> ORGANISM: Artificial Sequence
249 <220> FEATURE:
250 <223> OTHER INFORMATION: N-Terminal Inter-polypeptide linker
251 <400> SEQUENCE: 16
252 Val Ala Asp Glu Gly Gln Leu Arg Asp Tyr Leu Lys Arg Ala Ile Ala
253 1 5 10 15
254 Asp Ala Arg Asp Ala Arg Thr Arg Leu Arg Glu Val Glu Glu Ala
255 20 25 30
256 Arg
257 <210> SEQ ID NO: 17
258 <211> LENGTH: 30

```

## RAW SEQUENCE LISTING

DATE: 08/13/2002

PATENT APPLICATION: US/10/091,244

TIME: 13:46:44

Input Set : N:\jumbos\10091244DC.txt

Output Set: N:\CRF4\08132002\J091244.raw

174 <212> TYPE: PRT  
 175 <213> ORGANISM: Artificial Sequence  
 177 <220> FEATURE:  
 178 <223> OTHER INFORMATION: N-Terminal Inter-polypeptide linker  
 180 <400> SEQUENCE: 17  
 181 Met Ala Thr Asp Glu Lys Leu Leu Lys Tyr Leu Lys Arg Val Thr Ala  
 182 1 5 10 15  
 183 Glu Leu His Ser Leu Arg Lys Gln Gly Ala Arg His Ala Asp  
 184 20 25 30  
 187 <210> SEQ ID NO: 18  
 188 <211> LENGTH: 32  
 189 <212> TYPE: PRT  
 190 <213> ORGANISM: Artificial Sequence  
 192 <220> FEATURE:  
 193 <223> OTHER INFORMATION: N-Terminal Inter-polypeptide linker  
 195 <400> SEQUENCE: 18  
 196 Met Arg Glu Asp Gln Leu Leu Asp Ala Leu Arg Lys Ser Val Lys Glu  
 197 1 5 10 15  
 198 Asn Ala Arg Leu Arg Lys Ala Asn Thr Ser Leu Arg Ala Ala Met Asp  
 199 20 25 30  
 302 <210> SEQ ID NO: 19  
 303 <211> LENGTH: 33  
 304 <212> TYPE: PRT  
 305 <213> ORGANISM: Artificial Sequence  
 307 <220> FEATURE:  
 308 <223> OTHER INFORMATION: N-Terminal Inter-polypeptide linker  
 310 <400> SEQUENCE: 19  
 311 Met Pro Glu Gln Asp Lys Val Val Glu Tyr Leu Arg Trp Ala Thr Ala  
 312 1 5 10 15  
 313 Glu Leu His Thr Thr Arg Ala Lys Leu Glu Ala Leu Ala Ala Ala Asn  
 314 20 25 30  
 315 Thr  
 319 <210> SEQ ID NO: 20  
 320 <211> LENGTH: 31  
 321 <212> TYPE: PRT  
 322 <213> ORGANISM: Artificial Sequence  
 324 <220> FEATURE:  
 325 <223> OTHER INFORMATION: peptide *Unsequenced*  
 327 <400> SEQUENCE: 20  
 328 Met Thr Asp Ser Glu Lys Val Ala Glu Tyr Leu Arg Arg Ala Thr Leu  
 329 1 5 10 15  
 330 Asp Leu Arg Ala Ala Arg Gln Arg Ile Arg Glu Leu Glu Ser Asp  
 331 20 25 30  
 334 <210> SEQ ID NO: 21  
 335 <211> LENGTH: 25  
 336 <212> TYPE: DNA  
 337 <213> ORGANISM: Artificial Sequence  
 339 <220> FEATURE:  
 340 <223> OTHER INFORMATION: primer

VERIFICATION SUMMARY

PATENT APPLICATION: US/10/091,244

DATE: 08/13/2002

TIME: 13:46:45

Input Set : N:\jumbos\10091244DC.txt

Output Set: N:\CRF4\08132002\J091244.raw





OIPE

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/091,244

DATE: 08/08/2002

TIME: 10:55:37

Input Set : D:\30062-20046.txt

Output Set: N:\CRF3\08082002\J091244.raw

4 <110> APPLICANT: Gokhale, Rajesh  
 5       Isuji, Stuart  
 6       Khosla, Chaitan  
 8 <120> TITLE OF INVENTION: METHODS TO MEDIATE POLYKETIDE SYNTHASE  
 9       MODULE EFFECTIVENESS  
 11 <130> FILE REFERENCE: 300622004620  
 13 <140> CURRENT APPLICATION NUMBER: US 10/091,244  
 14 <141> CURRENT FILING DATE: 2002-03-04  
 16 <150> PRIOR APPLICATION NUMBER: 09/500,747  
 17 <151> PRIOR FILING DATE: 2000-02-04  
 19 <150> PRIOR APPLICATION NUMBER: 60/119,363  
 20 <151> PRIOR FILING DATE: 1999-02-04  
 22 <150> PRIOR APPLICATION NUMBER: 60/272,985  
 23 <151> PRIOR FILING DATE: 2001-03-02  
 25 <150> PRIOR APPLICATION NUMBER: 60/272,987  
 26 <151> PRIOR FILING DATE: 2001-03-02  
 28 <160> NUMBER OF SEQ ID NOS: 24  
 30 <170> SOFTWARE: FastSEQ for Windows Version 4.0

## ERRORED SEQUENCES

367 <210> SEQ ID NO: 24  
 368 <211> LENGTH: 24  
 369 <212> TYPE: DNA  
 370 <213> ORGANISM: Artificial Sequence  
 372 <220> FEATURE  
 373 <223> OTHER INFORMATION: primer  
 375 <400> SEQUENCE: 24  
 376 gaattcctac aggtctcttc cccc 24  
 E--> 378 1 g a a t t c c t a c a g g t c t c t c t c c c c

VERIFICATION SUMMARY

PATENT APPLICATION: US/10/091,244

DATE: 08/14/2002

TIME: 10:55:58

Input Set : D:\30062-20046.txt

Output Set: N:\CRF3\08082002\J091244.raw

L:378 M:204 E: No. of Bases conflict: this line has no nucleotides.